IST 686

**Final Examination Report**

Fall 2024

Prepare your final exam report by answering the questions in this document. In your answer for each question, make sure you write a narrative with complete sentences. Keep your answers concise and focused on the question asked but provide sufficient numeric and graphical detail that the staff member can create a comprehensive briefing for a legislator (see question 7 for specific points of interest). Make sure to include enough statistical information so that another analytics professional could review your work. You can assume that the staff member understands the concept of statistical significance and other basic concepts like mean, standard deviation, and correlation, so you do not need to define those. Be sure that you not only report what a test result was, but also what that result means substantively for the question you are answering.

You can choose to put important statistical values into a table for readability, or you can include the statistics within your narrative. Your report can include graphics created by R, keeping in mind that if you do include a graphic, you will have to provide some accompanying narrative text to explain what it is doing in your report. You can delete this instruction page and the prompts in italics but keep the section headings. Finally, be sure to proofread your final submission to eliminate typos, check grammar and to ensure that everything is included and readable.

I will assume that the statistical results in your answers refer to the corresponding section of your analysis notebook. If you decide you need to rerun an analysis, you can do so, but you will only receive credit for interpreting the new results. If you do rerun an analysis, please be sure to reference it and upload a knitted PDF with the analysis.

You may use an LLM such as ChatGPT to help prepare your answers. If you do so, you must disclose specifically how you used it in the final section of the report. You **may not** seek nor receive assistance, help, coaching, guidance, or support from any human except your instructor at any point during this exam. Seeking or obtaining improper assistance will result in a 0 for this exam. Your instructor will be available by email throughout the report writing period if you have questions, but don’t wait until the last minute!

Your responses will be graded on clarity; conciseness; inclusion and explanation of specific and appropriate statistical values; explanation of any included tabular material and the appropriate use of graphical displays when/if necessary. Bonus points will be awarded for work that goes above expectations.

# 1. Introductory paragraph

*In your own words, write about three sentences of introduction addressing the staff member in the state legislator’s office. Frame the problem/topic that your report addresses.*

**Vaccines are an important public health and safety measure nationwide, and it is important to use data to determine just how effective and widespread vaccine use is. This is an especially important issue for young children in schools, and this report will assess the level of vaccination among California school districts compared to the general United States population, as well as determine what factors lead to increased or decreased vaccination rates.**

# Descriptive Reporting

## 2. Descriptive Overview of U.S. Vaccinations

### a. How have U.S. vaccination rates varied over time? Are there significant trends or cyclical variation in U.S. vaccination rates?

**According to both the changepoint analysis and periodograms, there does not appear to be much seasonality or notable trends in vaccination rates for any disease in the time series. There appeared to be a large number of changepoints in each variable, suggesting that vaccination rates are very volatile, and don't follow structured uniform patterns. The periodograms also proved this, as it showed spikes close to the start of the x-axis, suggesting no long term seasonal trends.**

### b. What are the mean U.S. vaccination rates when including only recent years in the calculation of the mean?

### **Mean DTP Vaccination (1990-2017): 97.32143**

**Mean Polio Vaccination (1996-2017): 92.18182**

**Mean MMR Vaccination (1990-2017): 90.35714**

**Mean HepB Vaccination (2004-2017): 61.92857**

## 3. Descriptive Overview of California Vaccinations

### a. What are the mean levels of the four vaccination rate variables across districts?

### **Mean DTP Vaccination: 89.69102**

**Mean Polio Vaccination: 90.11111**

**Mean MMR Vaccination: 89.67884**

**Mean HepB Vaccination: 92.10502**

### b. Among districts, how are the vaccination rates for individual vaccines related? In other words, if there are students with one vaccine, are students likely to have all the others?

### **Based on the results of the correlation matrix, we can conclude that students in California school districts are likely to have all four vaccines listed in this dataset if they already have one, since every correlation coefficient is above 0.90.**

### c. How do these Californian vaccination levels compare to U.S. vaccination levels (recent years only)?

**After creating 95% confidence intervals for the difference in means among the four vaccination rates, we can conclude that DTP and Polio vaccination rates are significantly lower in California school districts than they are in the general US population, HepB vaccination rates are significantly higher in California school districts than they are in the general US population, and there is no significant difference in MMR vaccination rates between California school districts and the general US population.**

### d. Provide one or two sentences of your professional judgment about where California school districts stand with respect to vaccination rates in the larger context of the U.S.

**With the exception of Hepatitis B, California school districts seem to be behind the rest of the US in terms of vaccination rates, as they have significantly lower rates of two vaccinations and a lower mean vaccination rate of a third vaccination.**

# 4. Comparison of public and private schools

### a. What proportion of public schools and what proportion of private schools reported vaccination data?

**97.4% of public schools reported vaccination data, while only 84.7% of private schools reported vaccination data.**

### b. Was there any credible difference in reporting between public and private schools?

**Because the confidence interval of the difference in proportions is completely below zero, we can conclude that there is a significantly lower proportion of private schools that report vaccination data than public schools that report vaccination data.**

### c. Does the proportion of students with completely up-to-data vaccinations vary from county to county? Report significant details.

**Because the p-value for the ANOVA test between up-to-date vaccination rate and county is less than 0.05, we can reject the null hypothesis that there are no differences between counties in up-to-date vaccination rates. As a result, we can conclude that the proportion of students with completely up-to-date vaccinations varies from county to county.**

# 5. Inferential reporting about districts

### a. Which of the four predictor variables predicts the percentage of all enrolled students with completely up-to-date vaccines?

**According to the model, the most accurate predictors of up-to-date vaccination status are percentage of families living below the poverty line and total enrollment, with both being significant at the 0.01 level. To be specific, enrollment is the most significant variable, as its p-value of 0.00207 is the lowest among all predictors.**

### b. Using any set or combination of predictors that you want to use, what combination gives the best R-squared in predicting the percentage of all enrolled students with belief exceptions?

**According to the correlation matrix of all variables in the data set, the best predictors of proportion of students with belief exemptions are all vaccination rate variables. To eliminate any risk of multicollinearity, I will leave out PctUpToDate, since it is accounted for by the individual vaccination rate factors.**

### c. In predicting the percentage of all enrolled students with completely up-to-date vaccines, is there an interaction between PctFamilyPoverty and Enrolled? If so, interpret the interaction term.

**The interaction term between PctFamilyPoverty and Enrollment is statistically significant at the 0.05 level, showing that there is likely an interaction between the two variables. because it is negative, we can conclude that students are less likely to have up-to-date vaccination status if they are in a school district with a higher percentage of families living below the poverty line and a larger number of enrolled students.**

### d. Which, if any, of the four predictor variables predict whether or not a district’s reporting was complete?

**Total schools and enrollment seem to be the only significant predictors.**

# 6. Concluding Paragraph

*Describe your conclusions, based on the foregoing analyses. The staff member in the state legislator’s office is interested to know how to allocate financial assistance to school districts to improve both their vaccination rates and their reporting compliance. Make sure you have at least one sentence that makes a recommendation about improving vaccination rates and at least one that makes a recommendation about improving reporting rates. Finally, say what further analyses might be helpful to answer these questions and any additional data you would like to have.*

**Since poverty rates have a high correlation with vaccination rates, it would be wise to allocate financial assistance to lower income districts, to ensure families are financially able to get their kids vaccinated. In order to improve the rate at which schools report vaccinations, it seems to be most important to increase accessibility to vaccines, as schools with lower vaccination rates overall tended to report their data less.**

# LLM disclosure

If you used an LLM to prepare any answers to this exam, disclose so here. For each question, paste in the prompt you gave and explain how you used the response you received.